1645

TRANSMITTAL LETTER			DOCKET NUMBER: P-IS 4588
SERIAL NO: 09/823,850	FILING DATE: March 30, 2001	EXAMINER: Not Yet Known	GROUP ART UNIT: 1645
INVENTION: METHODS FOR DETERMINING THE TRUE SIGNAL OF AN ANALYTE			

TO COMMISSIONER FOR PATENTS

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Washington, D.C., 20231 on October 8, 2002.

By: David A. Gay, Reg. No. 39,200

October 8, 2002
Date of Signature

Transmitted herewith are the following documents in connection with the above-identified application:

1. Request for Corrected Patent Application Publication (in duplicate).
2. Exhibits A.
3. Return postcard.

___ Please charge my Deposit Account No. 03-0370 the amount of $_____. A duplicate copy of this sheet is enclosed.

_X_ The Commissioner is hereby authorized to charge payment of any fees associated with this communication or credit any overpayment to Deposit Account No. 03-0370. A duplicate copy of this sheet is enclosed.

_X_ The Commissioner is hereby authorized to charge to Deposit Account No. 03-0370 any fees under 37 CFR 1.17 which may be required under 37 CFR 1.136(a)(3) for an extension of time in any concurrent or future reply requiring a petition for extension of time. A duplicate copy of this sheet is enclosed.

Respectfully submitted,

David A. Gay
Registration No. 39,200
CAMPBELL & FLORES LLP
4370 La Jolla Village Drive
7th Floor
San Diego, California 92122
858-535-9001
USPTO CUSTOMER NO. 23601

PATENT
Our Docket: P-IS 4588

#8
Plunkett
10/22/02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:)
Ideker et al.)
)
Serial No.: 09/823,850)
)
Filed: March 30, 2001)
)
For: METHODS FOR DETERMINING THE)
TRUE SIGNAL OF AN ANALYTE)
)

Group Art Unit: 1645

Examiner: Not Yet Known

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Washington, D.C., 20231 on October 8, 2002.

By: _____
David A. Gay, Reg. No. 39,200

October 8, 2002
Date of Signature

Commissioner for Patents
Washington, D.C. 20231

REQUEST FOR CORRECTED PATENT APPLICATION PUBLICATION

Pursuant to 37 C.F.R. § 1.221(b), the publication of the above-identified application, publication No. US-2002-0107640-A1, published August 8, 2002, is respectfully requested to be corrected.

On page 11, paragraph [0092], please delete "*P*" in the first formula and replace with --p--.

On page 15, claim 17, please delete "gorup" and replace with --group--.

On page 16, claim 52, paragraph (d), please delete "signal p" and replace with --signal μ--.

On page 17, claim 66, please delete "form" and replace with --from--.

Submitted herewith is a copy of pages 11, 15 through 17, of the publication as it was published on August 8, 2002, with corrections noted thereon (Exhibit A).

Accordingly, Applicants request that this errors be corrected in the USPTO's electronic copy of the Specification and that the USPTO publish a corrected patent application publication.

No fee is deemed necessary to file this Request. If any fee is required, authorization is hereby given to charge the amount to Deposit Account No. 03-0370. A duplicate copy of this sheet is enclosed for this purpose.

Respectfully submitted,

October 8, 2002
Date

David A. Gay
Registration No. 39,200
Telephone No.: (858) 535-9001
Facsimile No.: (858) 535-8949

CAMPBELL & FLORES LLP
4370 La Jolla Village Drive
Suite 700
San Diego, California 92122
USPTO CUSTOMER NO. 23601

[0089] where (μ_{xi},μ_{yi}) is the pair of true mean intensities for gene i. For each i and j, the multiplicative errors ϵ_{xij} and ϵ_{yij}, are drawn from a bivariate normal distribution with means 0, standard deviations $\sigma_{\epsilon x}$ and $\sigma_{\epsilon y}$, and correlation ρ_ϵ. The additive errors δ_{xij} and δ_{yij}, are distributed analogously, with parameters $\sigma_{\delta x}$, $\sigma_{\delta y}$ and ρ_δ. Thus, multiplicative and additive errors are independent of one another but can each be highly correlated between x and y; in practice ρ_ϵ is large and ρ_δ is small. While x_{ij} and y_{ij} can be negative if the foreground is less than the estimated background for a spot, the true intensities μ_{xi} and μ_{yi} must be non-negative. Consequently, the samples (x_{ij} and y_{ij}) are described by a bivariate normal probability density function p with parameters μ_{xi} and μ_{yi} σ_{xi}, σ_{yi} and $\rho_{xi,yi}$, where:

$$\sigma_{xi} = \sqrt{\mu_{xi}^2\sigma_{\epsilon x}^2 + \sigma_{\delta x}^2}$$

$$\sigma_{yi} = \sqrt{\mu_{yi}^2\sigma_{\epsilon y}^2 + \sigma_{\delta y}^2}$$

$$\rho_{xi,yi} = \frac{\mu_{xi}\mu_{yi}\rho_\epsilon\sigma_{\epsilon x}\sigma_{\epsilon y} + \rho_\delta\sigma_{\delta x}\sigma_{\delta y}}{\sigma_{xi}\sigma_{yi}}$$

[0090] The model depends on six gene-independent parameters $\beta=(\sigma_{\epsilon x}, \sigma_{\epsilon y}, \rho_\epsilon, \sigma_{\delta x}, \sigma_{\delta y}, \rho_\delta)$ and a mean pair per gene, $\mu=[(\mu_{x1},\mu_{y1}), (\mu_{x2},\mu_{y2}), \ldots, (\mu_{xN},\mu_{yN})]$ for a total of 2N+6 parameters. The probability density function for gene i is $p=p(x_{ij}, y_{ij}|\beta, \mu_{xi}, \mu_{yi})$.

[0091] Parameter Estimation by Maximum Likelihood

[0092] Since β and μ are generally unknown, they can be estimated by using a maximum likelihood estimation (MLE) as described by Kendall and Stuart, *The Advanced Theory of Statistics*, Volume 2 (4th ed., Macmillan Publishing Co., New York, N.Y., 1979), which is incorporated herein by reference. Likelihood functions, for gene i and over all genes, are respectively defined as:

$$L_i(\beta, \mu_{xi}, \mu_{yi}) = \prod_{j=1}^{M} p(x_{ij}, y_{ij} / \beta, \mu_{xi}, \mu_{yi})$$

$$L(\beta, \mu) = \prod_{i=1}^{N} L_i(\beta, \mu_{xi}, \mu_{yi})$$

[0093] The MLE parameter values maximizing L, designated β and μ, are estimates for the true parameters of the underlying statistical model. In general, these values can be found using standard optimization procedures as described by Press et al., *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed., Cambridge University Press, Cambridge, Mass.). Because N can be large β and μ, can be determined by optimizing subsets of parameters in separate stages:

[0094] (1) choose initial values for μ,

[0095] (2) select β to maximize L given current values of μ,

[0096] (3) for i=1, . . . , N: select (μ_{xi},μ_{yi}) to maximize L_i, given current values of β, and

[0097] (4) repeat (2) and (3) until β, μ have converged.

[0098] All stages of the optimization were performed using the procedure fmincon provided by Matlab and described by Coleman et al., *Matlab Optimization Toolbox User's Guide* (3rd ed., Mathworks, Inc., Natick, Mass., 1999), which was incorporated herein by reference. The optimization was also implemented in C code, which produces comparable optimal parameters in substantially less execution time (less than 10 minutes on a Pentium III 500 for N=6000, M=4, as compared with 4-5 hours for the Matlab implementation). In both cases, all parameters converged within 250 iterations of stages (2) and (3) and are insensitive to initial choices for β and μ.

[0099] Significance Testing using Likelihood Ratios

[0100] After the parameters have been determined for a given set of observations, it is of immediate interest to use the model to identify mean intensity pairs which are significantly unequal such that $\mu_{xi}\neq\mu_{yi}$, representing genes that are differentially expressed between the two cell populations. For each gene i, the generalized likelihood ratio test (GLRT) (Kendall and Stuart 1979) statistic λ_i is computed according to:

$$\lambda_i = -2\ln\left(\frac{\max_{\mu} L_i(\beta, \mu, \mu)}{\max_{\mu_x,\mu_y} L_i(\beta, \mu_x, \mu_y)}\right)$$

[0101] Two maximizations are performed: in the numerator, the constraint $\mu_x=\mu_y=\mu$ is imposed, while in the denominator the optimization is unconstrained. Under the null hypothesis that $\mu_x=\mu_y$, β remains a consistent estimator when the constraint is imposed.

[0102] In the case that $\mu_{xi}=\mu_{yi}$, μ_i follows (asymptotically in M and N) a χ^2 distribution with 1 degree of freedom (DOF), whereas if $\mu_{xi}\neq\mu_{yi}$, the value of λ_i is expected to be larger than would be obtained from random sampling of this distribution. To select differentially-expressed genes with a selection error of α, the false positive or Type-I error rate, one would first determine the critical value λ_c, for which the χ^2 cumulative probability distribution is equal to $1-\alpha$, then select the set of all genes i for which λ_i is in the critical region $\lambda_i>\lambda_c$. The particular choice of a depends on the number of genes on the array and the selection error which the individual investigator is willing to tolerate.

EXAMPLE II

Identification of Genes Differentially-Expressed in Response to Galactose Stimulation of Yeast Cells

[0103] This example describes application of the mathematical model of the variability observed over repeated observations of intensities for genes represented on a DNA microarray to the identification of genes differentially-expressed in response to galactose stimulation.

[0104] Assembly of the Microarray

[0105] In order to explore the performance of the test for differentially-expressed genes as shown in Example I, *Saccharomyces cerevisiae* cultures growing in the absence of galactose (YPR media) were compared to those growing in galactose-stimulating conditions (YPRG) using a DNA

2. The method of claim 1, further comprising selecting a mean signal μ that provides a maximum probability of likelihood given said observed signal.

3. The method of claim 1, wherein said additive and multiplicative errors are independent with respect to each other.

4. The method of claim 1, wherein said observed signal and said mean signal further comprises the relationship;

$$x_{ij}=\mu_{xi}+\mu_{xi}\epsilon_{xij}+\delta_{xij},$$

where each measurement j=1, . . . , M, each analyte i=1, . . . , N, and where x_{ij} is the observed signal and μ_{xi} is the mean signal.

5. The method of claim 1, wherein said additive and multiplicative errors further comprise a univariate distribution.

6. The method of claim 5, wherein said univariate distribution is a parametric distribution.

7. The method of claim 6, wherein said parametric distribution is a univariate normal distribution.

8. The method of claim 7, wherein said univariate normal distribution and said system parameter further comprise a multiplicative error term consisting of a normal distribution having standard deviation with respect to a signal mean ($\sigma_{\epsilon x}$) and an additive error term consisting of a normal distribution having standard deviation with respect to a signal mean ($\sigma_{\delta x}$).

9. The method of claim 6, wherein said parametric distribution is a t-distribution.

10. The method of claim 6, wherein said parametric distribution is a gamma distribution.

11. The method of claim 1, wherein said mean signal and system parameter are determined at the same time.

12. The method of claim 1, wherein said system parameter is determined before said mean signal is determined.

13. The method of claim 12, wherein said predetermined system parameter is used to determine said mean signal.

14. The method of claim 1, wherein said enhanced values for said probability likelihood of said observed signals are produced one or more times until said mean signal and said system parameter converge.

15. The method of claim 1, wherein said mean signal and said system parameter are determined by a method selected from the group consisting of maximum likelihood estimation (MLE), Quasi-Maximum Likelihood and Generalized Method of Moments.

16. The method of claim 1, wherein determining said mean signal and said system parameter further comprises a non-linear optimization algorithm.

17. The method of claim 16, wherein said optimization algorithm is selected from the ~~group~~ consisting of Gradient Descent, Newton-Raphson and Simulated Annealing.

18. A method of determining a true signal of an analyte, comprising:

(a) obtaining an observed signal x for one or more analytes;

(b) providing a mean signal (μ) and a system parameter (β) for said analyte;

(c) computing a probability likelihood of said observed signal, said observed signal being related to said mean signal by an additive error (δ) and a multiplicative error (ϵ), where said system parameter specifies properties of said additive error and said multiplicative error, and

(d) selecting a mean signal μ and a system parameter (β) that provides a maximum probability likelihood of occurrence given said observed signal.

19. The method of claim 18, wherein said additive and multiplicative errors are independent with respect to each other.

20. The method of claim 18, wherein said observed signal and said mean signal further comprises the relationship:

$$x_{ij}=\mu_{xi}+\mu_{xi}\epsilon_{xij}+\delta_{xij},$$

where each measurement j=1, . . . , N, each analyte i=1, . . . , N, and where x_{ij} is the observed signal and μ_{xi} is the mean signal.

21. The method of claim 18, wherein said additive and multiplicative errors further comprise a univariate distribution.

22. The method of claim 1, wherein said univariate distribution is a parametric distribution.

23. The method of claim 22, wherein said parametric distribution is a univariate normal distribution.

24. The method of claim 23, wherein said univariate normal distribution and said system parameter further comprise a multiplicative error term consisting of a normal distribution having standard deviation with respect to a signal mean ($\sigma_{\epsilon x}$), and an additive error term consisting of a normal distribution having standard deviation with respect to a signal mean ($\sigma_{\delta x}$).

25. The method of claim 22, wherein said parametric distribution is a t-distribution.

26. The method of claim 22, wherein said parametric distribution is a gamma distribution.

27. The method of claim 18, wherein said mean signal and system parameter are selected at the same time.

28. The method of claim 18, wherein said system parameter is selected before said mean signal is determined.

29. The method of claim 28, wherein said preselected system parameter is used to select said mean signal.

30. The method of claim 18, further comprising computing said probability likelihood one or more times until said mean signal and said system parameter converge.

31. The method of claim 18, wherein said mean signal and said system parameter are determined by a method selected from the group consisting of maximum likelihood estimation (MLE), Quasi-Maximum Likelihood and Generalized Method of Moments.

32. The method of claim 18, wherein selecting said mean signal and said system parameter further comprises a non-linear optimization algorithm.

33. The method of claim 32, wherein said optimization algorithm is selected from the group consisting of Gradient Descent, Newton-Raphson and Simulated Annealing.

34. A method of determining relative amounts of an analyte between samples, comprising:

(a) measuring observed signals x and y for an analyte within two or more sample pairs, and

(b) determining a mean signal pair per analyte (μ) and a system parameter (β) for each sample pair that produce enhanced values for a probability likelihood of said observed signals, said observed signals being related to said mean signals by an additive error (δ) and a multiplicative error (ϵ), wherein said system parameter specifies properties of said additive error (δ) and said multiplicative error (ϵ).

35. The method of claim 34, further comprising selecting a mean signal p that provides a maximum probability of occurrence given said observed signals.

36. The method of claim 34, wherein said additive and multiplicative errors are independent with respect to each other.

37. The method of claim 34, wherein said observed signals and said mean signal pair per analyte within said sample pairs further comprise the relationship:

$$x_{ij}=\mu_{xi}+\mu_{xi}\epsilon_{xij}+\delta_{xij}, \text{ and}$$

$$y_{ij}=\mu_{yi}+\mu_{yi}\epsilon_{yij}+\delta_{yij}$$

where each measurement j equals 1 through M and each analyte i equals 1 through N; where x_{ij} and y_{ij} are the observed signals, and where μ_{xi} and μ_{yi} are the mean signals.

38. The method of claim 34, wherein said additive and multiplicative errors further comprise a bivariate distribution.

39. The method of claim 38, wherein said bivariate distribution is a parametric distribution.

40. The method of claim 38, wherein said parametric distribution is a bivariate normal distribution.

41. The method of claim 40, wherein said bivariate normal distribution and said system parameter further comprises a multiplicative error term consisting of a standard deviation with respect to a mean of signal x ($\sigma_{\epsilon x}$), a standard deviation with respect to a mean of signal y ($\sigma_{\epsilon y}$) and a correlation between signals x and y (ρ_ϵ), and an additive error term consisting of a standard deviation with respect to a mean of signal x ($\sigma_{\delta x}$), a standard deviation with respect to a mean of signal y ($\sigma_{\delta x}$) and a correlation between signals x and y (ρ_δ).

42. The method of claim 39, wherein said parametric distribution is a t-distribution.

43. The method of claim 39, wherein said parametric distribution is a bivariate gamma distribution.

44. The method of claim 34, wherein said mean signal pair per analyte and system parameter are determined at the same time.

45. The method of claim 34, wherein said system parameter is determined before said mean signal pair per analyte is determined.

46. The method of claim 45, wherein said predetermined system parameter is used to determine said mean signal pair per analyte.

47. The method of claim 34, wherein said enhanced values for said probability likelihood of said observed signals are produced one or more times until said mean signal pair per analyte and said system parameter converge.

48. The method of claim 34, wherein determining said mean signal pair per analyte and said system parameter further comprises a non-linear optimization algorithm.

49. The method of claim 48, wherein said optimization algorithm is selected from the group consisting of Gradient Descent, Newton-Raphson and Simulated Annealing.

50. The method of claim 34, further comprising identifying significantly unequal mean signal pairs per analyte by a statistical difference indicator.

51. The method of claim 50, wherein said difference indicator further comprises a generalized likelihood ratio test statistic (λ).

52. A method of determining relative amounts of an analyte between samples, comprising:

(a) obtaining observed signals x and y for an analyte within two or more sample pairs;

(b) providing a mean signal pair per analyte (μ) and a system parameter (β) for each sample pair;

(c) computing a probability likelihood of said observed signals, said observed signals being related to said mean signal by an additive error (δ) and a multiplicative error (ϵ), where said system parameter specifies the properties of said additive error and said multiplicative error, and

(d) selecting a mean signal $\vec{p}$ and a system parameter (β) that provides a maximum probability likelihood of occurrence given said observed signals.

53. The method of claim 52, wherein said additive and multiplicative errors are independent with respect to each other.

54. The method of claim 52, wherein said observed signals and said mean signal pair per analyte within said sample pairs further comprise the relationship:

$$x_{ij}=\mu_{xi}+\mu_{xi}\epsilon_{xij}+\delta_{xij}, \text{ and}$$

$$y_{ij}=\mu_{yi}+\mu_{yi}\epsilon_{yij}+\delta_{yij}$$

where each measurement j equals 1 through M and each analyte i equals 1 through N; where x_{ij} and y_{ij} are the observed signals, and where μ_{xi} and μ_{yi} are the mean signals.

55. The method of claim 52, wherein said additive and multiplicative errors further comprise a bivariate distribution.

56. The method of claim 55, wherein said bivariate distribution is a parametric distribution.

57. The method of claim 56, wherein said parametric distribution is a bivariate normal distribution.

58. The method of claim 57, wherein said bivariate normal distribution and said system parameter further comprise a multiplicative error term consisting of a standard deviation with respect to a mean of signal x ($\sigma_{\epsilon x}$), a standard deviation with respect to a mean of signal y ($\sigma_{\epsilon y}$) and a correlation between signals x and y (ρ_ϵ), and an additive error term consisting of a standard deviation with respect to a mean of signal x ($\sigma_{\delta x}$), a standard deviation with respect to a mean of signal y ($\sigma_{\delta y}$) and a correlation between signals x and y (ρ_δ).

59. The method of claim 56, wherein said parametric distribution is a t-distribution.

60. The method of claim 56, wherein said mean signal pair per analyte and system parameter are determined at the same time.

61. The method of claim 52, wherein said system parameter is determined before said mean signal pair per analyte is determined.

62. The method of claim 61, wherein said predetermined system parameter is used to determine said mean signal pair per analyte.

63. The method of claim 52, further comprising computing said probability likelihood of said observed signals one or more times until said mean signal pair per analyte and said system parameter converge.

64. The method of claim 52, wherein said mean signal pair per analyte and said system parameter are determined by a method selected from the group consisting of maximum likelihood estimation (MLE), Quasi-Maximum Likelihood and Generalized Method of Moments.

65. The method of claim 52, wherein selecting said mean signal pair per analyte and said system parameter further comprises a non-linear optimization algorithm.

66. The method of claim 65, wherein said optimization algorithm is selected from the group consisting of Gradient Descent, Newton-Raphson and Simulated Annealing.

67. The method of claim 52, further comprising identifying said mean signal pair per analyte that are significantly unequal using a difference indicator.

68. The method of claim 67, wherein said difference indicator further comprises a generalized likelihood ratio test statistic (λ).

69. The method of claim 67, further comprising selecting two or more mean signal pairs per analyte having a difference indicator greater than that corresponding to a false positive error rate.

70. The method of claim 52, wherein said analyte is a nucleic acid or polypeptide.

71. A method of determining relative amounts of analytes between samples, comprising:

(a) obtaining observed signals x and y for a plurality of immobilized analytes within two or more sample pairs;

(b) determining a mean signal pair per analyte (μ) and a system parameter (β) for each sample pair that provides a maximum probability likelihood of occurrence given said observed signals, said observed signals being related to said mean signal by an additive error (δ) and a multiplicative error (ϵ), where said system parameter specifies the properties of said additive error and said multiplicative error, and

(c) identifying one or more mean signal pairs per analyte that is significantly unequal.

72. The method of claim 71, wherein said additive and multiplicative errors are independent with respect to each other.

73. The method of claim 71, wherein said observed signals and said mean signal pair per analyte within said sample pairs further comprise the relationship:

$$x_{ij}=\mu_{xi}+\mu_{xi}\epsilon_{xij}+\delta_{xij}, \text{ and}$$

$$y_{ij}=\mu_{yi}+\mu_{yi}\epsilon_{yij}+\delta_{yij}$$

where each measurement j equals 1 through M and each analyte i equals 1 through N; where x_{ij} and y_{ij} are the observed signals, and where μ_{xi} and μ_{yi} are the mean signals.

74. The method of claim 71, wherein said one or more mean signal pairs per analyte are identified as significantly unequal by using a difference indicator.

75. The method of claim 74, wherein said difference indicator further comprises a generalized likelihood ratio test statistic (λ).

76. The method of claim 74, further comprising selecting two or more mean signal pairs per analyte having a difference indicator greater than that corresponding to a false positive error rate.

77. The method of claim 71, wherein said analyte is a nucleic acid or polypeptide.

78. The method of claim 71, wherein said plurality of analytes further comprises about 1,000 or more different analytes.

79. The method of claim 71, wherein said plurality of analytes further comprises about 10,000 or more different analytes.

80. The method of claim 71, wherein said plurality of analytes further comprises about 30,000 or more different analytes.

81. The method of claim 71, further comprising analytes mobilized on a microarray.

82. The method of claim 71, further comprising the steps of:

(a) obtaining one or more reference signals, and

(b) determining a mean signal pair (μ) and a system parameter (β) for a sample pair comprising said observed signal x or y and said reference signal that provides a maximum probability likelihood of occurrence given said reference and observed signals, said reference and observed signals being related to said mean signal by an additive error (δ) and a multiplicative error (ϵ), wherein said system parameter specifies the properties of said additive error and said multiplicative error.

83. A method of determining relative amounts of an analyte between samples, comprising:

(a) obtaining a reference signal;

(b) obtaining observed signals x and y for an analyte within two or more sample pairs;

(c) determining system parameters (β_1, β_2) for a sample pair comprising said observed signals x or y and said reference signal that provide a probability likelihood of said occurrence given said observed and reference signals, said observed and reference signals being related to said mean signal by an additive error (δ) and a multiplicative error (ϵ), where said system parameter specifies the properties of said additive error and said multiplicative error;

(d) determining mean signal pairs (μ_1, μ_2) for said sample pair comprising maximizing a product of terms for said probability likelihood of said sample pair of observed signals x or y and said reference signal for said analyte, and

(e) selecting a mean signal μ_x or μ_y that provides a maximum probability likelihood of occurrence given said observed signals and system parameters β_1 and β_2.

84. The method of claim 83, wherein said mean signal pairs (μ_1, μ_2) are determined using β_1 and β_2 obtained from step (c).

85. A method of determining relative amounts of an analyte between samples, comprising:

(a) measuring observed signals x, y and z for an analyte within two or more sample sets, and

(b) determining a mean signal set per analyte (μ) and a system parameter (β) for each sample set that produce enhanced values for a probability likelihood for said observed signals, said observed signals being related to mean signals by an additive error (δ) and a multiplicative error (ϵ).

* * * * *